

О МЕТОДИКЕ СБОРА ЛЕКСИКИ ПОДЪЯЗЫКОВ НАУКИ И ТЕХНИКИ
И СОСТАВЛЕНИЯ ЧАСТОТНЫХ СЛОВАРЕЙ
(на материале немецкого подъязыка
экономики промышленности)

Основной целью обучения иностранному языку в неязыковом вузе является научить "читать оригинальную литературу по специальности для извлечения необходимой информации" (1), для чего необходим строгий отбор лексики по определенному профилю, ибо программа неязыкового вуза весьма ограничена в учебных часах.

Под "отбором учебного материала" подразумевается отбор языковых средств подъязыков различных рангов (2) и их реализация в речевом материале – системе тренировочных и речевых упражнений, учебников и учебных пособий, словарей-минимумов и т.п. (3). Основой словаря-минимума должен стать частотный словарь. О необходимости создания словарей по отдельным отраслям говорит А.К.Рейцак (4). Однако словарь, составленный на основе статистического метода, дает неполное представление о языке*. Список, составленный только на основании показаний частотности, не имеет достаточного нормативного значения. (5). *Vocabulaire de disponibilité* может быть составлен только по какой-то определенной теме, для определенного контекста. Этот контекст должен быть всегда строго определенным (6). Однако словарь не составляется, а "творится", как образно заметил Л.В.Щерба, указывая на творческий подход к работе по созданию словарей. Он говорил, что работа по созданию словарей "должна иметь научный характер и никак не состоять в механическом составлении каких-то готовых элементов" (7).

Под "творческим подходом" понимается научно обоснованный отбор наиболее употребительных языковых средств.

* Еще в 1920 г. американский лингвист Кенигстон предложил сопровождать каждое слово словаря, помимо частоты, индексом ранга, т.е. указанием на количество текстов, где данное слово встретилось. Это предложение было принято и вошло в практику частотных словарей.

Хотя проблеме отбора наиболее употребительной лексики посвящено в настоящее время немало работ (8), тем не менее она требует дальнейшего уточнения и разработки.

Известно, что изучение тех или иных языковых средств статистическим методом с последующей качественной оценкой статистически обработанного материала является необходимым условием для объективного отбора учебного материала – языкового и речевого.

В настоящее время такие критерии отбора лексики, как частота (абсолютная, относительная) и распространенность, т.е. равномерность распределения слова в обследуемых выборках, можно считать общепринятыми. Но традиционные критерии уже не могут в полной мере удовлетворить запросы лингвистов-статистиков, методистов и представителей других специальностей, использующих частотные словари. Вопрос о существенном улучшении техники составления частотных словарей, о расширении их лексикологического и статистического аппарата до сих пор остается нерешенным.

Все еще ведутся дискуссии о том, следует ли эти критерии рассматривать как основные или ставить их в один ряд с такими критериями, как словобразовательная способность, семантическая ценность, стилистическая ограниченность, сочетаемость слова и др.

Дискуссионными остаются понимание термина "употребительность слова" и объективная оценка этого понятия, которым нередко пользуются авторы частотных словарей и лексических минимумов, как правило, в виде обобщения принятых в них количественных и качественных критериев отбора лексики.

Интересная попытка ответить на этот вопрос сделана в работе М.В.Григорьевой (9). Творчески используя и углубляя результаты исследований советских и зарубежных лингвистов и методистов, в частности В.М.Андрюшенко, Е.Д.Розанова, Э.А.Штейнфельдт, А.Луиань (A.Juilland) и др., она рассматривает "употребительность" как самостоятельный принцип отбора лексики, т.е. в ряду других принципов, таких как критерий, количественно определяющий степень участия языковых средств, в данном случае лексики, в процессе реализации в соответствующем подязыке (9).

Применяя вероятностно-статистический метод, М.В.Григорьева дает количественное выражение "употребительности" в виде комплексного показателя четырех факторов: абсолютной частоты, распрост-

раненности, величины объема выборки и количества текстов (подвыборок) в составе выборки (9).

Из частотных словарей, предназначенных для учебных целей, наибольшую ценность представляют двуязычные частотные словари, в которых наряду с частотой учитываются другие количественные, а также качественные показатели. Необходимо, однако, отметить, что иноязычно-русских частотных словарей для неязыковых вузов все еще мало. Объясняется это отсутствием строго обоснованных критериев отбора и трудоемкостью отбора лексики (10, 11).

Для экономических вузов издано пока два частотных словаря: "Англо-русский экономико-статистический словарь" В.В.Морозенко (М., 1974) и "Частотный англо-русский словарь общэкономических терминов" Л.М.Пашина под редакцией проф. В.В.Морозенко (Куйбышев, 1974). Их достоинством является то, что каждая лексема сопровождается тремя показателями – частоты, распространенности и тематической ценности. Эти словари являются надежной основой для составления различного назначения словарей минимумов и написания учебников и учебных пособий экономико-статистического и экономического профиля. Интерес представляет также "Лексический минимум по английскому языку для технических вузов" В.В.Морозенко и И.Ф.Турук (М., 1975). Достоинством этого пособия является то, что оно, кроме общепринятых принципов частотности и распространенности, вводит новые критерии: 1) многофункциональность слова, под которой подразумевается его синтаксическая функция в речи; 2) словособразовательная способность слова; 3) семантическая ценность. Все эти дополняющие друг друга критерии получают в пособии количественное выражение. Ценно и то, что составители выделяют 20 тематических групп.

Обзор словарей показывает, что научно обоснованный отбор объективно необходимого учебного лексического минимума немалым без статистической характеристики лексем. Вместе с тем потребуются дальнейшие исследования, чтобы определить, какие именно количественные и качественные критерии следует считать решающими для отбора каждой лексической единицы в качестве учебной. Впрочем, это сознают и сами составители рассмотренных выше словарей, не претендуя на полноту реализации всех возможностей, которыми располагает методика при разрешении проблемы отбора лексики для обучения иностранному языку (12).

В связи с расширением работ по составлению различного типа отраслевых частотных словарей, увеличением массивов текстов, усложнением показателей и естественным желанием составителей словарей скорее использовать полученные результаты в учебном процессе статистическая обработка данных ручным методом все чаще уступает место ЭВМ. Применение ЭВМ для составления частотных словарей является по сравнению с ручным методом обработки не только менее громоздкой и утомительной процедурой, но и дает, что весьма важно, немалый выигрыш во времени при условии использования более совершенных вычислительных систем (13) и наличия тщательно разработанной методики (14).

Что касается непосредственно методики получения словарей с помощью ЭВМ, то и такая методика детально разработана целым рядом исследователей (15), в частности описано получение частотного словаря на машине "Минск-32" из объема текста в 400 000 слов — употребления по родственным немецким текстам (16). Не останавливаясь здесь на проблеме кодирования вводимых в машину словоформ и на других технических вопросах, мы ставим своей целью только уточнить методику отбора с помощью ЭВМ наиболее употребительной лексики научно-технического подъязыка на примере немецкого подъязыка экономики промышленности (ЭП).

Процесс отбора наиболее употребительной лексики можно разбить на следующие этапы:

- 1) определение объема выборки;
- 2) способы формирования подвыборок;
- 3) определение тематики подвыборок;
- 4) подготовка лингвистической информации для обработки на ЭВМ "Минск-2";
- 5) составление программы для ЭВМ "Минск-2";
- 6) ввод в машину; контроль;
- 7) получение частотных списков словоформ;
- 8) получение алфавитного частотного словаря.

Вопросы стратификации выборочных совокупностей текстов не имеют пока однозначного решения. Неясно, какое количество подвыборок можно считать достаточно большим, с тем чтобы иметь возможность объективно судить о закономерностях статистического распределения лексики. Неясны также критерии формирования и нормы подвыборок.

По мнению некоторых исследователей, объем выборки должен составить не менее 100-150 тысяч словоформ для получения более достоверных данных (17). М.В.Григорьева предлагает остановиться на выборке, равной 140 тысячам словоформ (9).

Опираясь на опыт лингвистов, мы остановились на выборке в 140 тысяч словоупотреблений.

Критерии формирования подвыборок тоже не едины. Одни исследователи считают, что наиболее эффективными являются подвыборки одинаковой длины: 100-500 словоформ (В.М.Андрющенко), 1000 (П.М.Алексеев), 2500 (Е.Д.Розанов), для других же важна не длина текста, а законченная мысль каждого текста, т.е. подвыборки разной длины (Н.Г.Вишнякова, И.Ф.Динабург; Э.А.Штейнфельдт, Л.М.Пашин). Интересен способ формирования подвыборок у М.Е.Кашириной, которая берет подвыборки объемом в 1000 словоформ, но, учитывая длину законченного текста, берет по 1000 словоформ в начале, середине и конце текста (19).

Принимая во внимание опыт указанных лингвистов и то, что при равных подвыборках легче производить все математические подсчеты, мы остановились на методике М.Е.Кашириной. Таким образом, мы имеем 140 подвыборок объемом по 1000 словоформ каждая.

Подвыборки взяты из немецкой оригинальной литературы с учетом жанрово-тематической дифференциации:

1) монографии и диссертации, ставящие своей целью передачу новых сведений и описание их применения на практике, - 10%;

2) учебники, учебные пособия, объединяемые научно-популярной направленностью, - 40%;

3) статьи, рефераты в научных журналах - 30%, брошюры, справочники, цель которых обобщить имеющиеся сведения и опыт, предложить обзор уже достигнутого, - 10%;

4) газеты (прежде всего экономические), в которых передаются и обобщаются новые сведения и опыт достигнутого, - 10%.

Все это жанровое многообразие скреплено единством функции, подчинившей себе все остальное: закрепить, сохранить, переработать и передать разнобразную научно-техническую информацию.

Что же касается тематической дифференциации, то вся литература по ЭП подбиралась по 9 основным разделам:

Тематика	Количество текстов	Объем, %
Организация управления производством	14	10
Планирование производства и его эффективность	28	20
Концентрация, кооперирование, специализация и комбинирование	14	10
Основные производственные фонды и капиталовложения	17	12,1
Материально-техническое снабжение и оборотные средства	14	10
Труд	14	14
Себестоимость, цены, финансы, учет, хозрасчет	15	10,7
Производственно-технический прогресс и применение ЭВМ в управлении производством	12	8,6
Производственно-технологические вопросы	12	8,6
Итого:	140	100

Среднее процентное соотношение текстов того или иного раздела определялось нами в соответствии с объемом подобных же разделов в учебниках и справочниках (20).

Принимая во внимание принципы отбора, предложенные Н.Г.Вишняковой, Е.Д.Розановым, М.В.Григорьевой (21), мы предлагаем принимать во внимание статистическую ограниченность.

Под принципом статистической ограниченности понимаем:

1) абсолютную частоту (т.е. сколько раз слово встретилось в 140 подвыборках); 2) распространенность (в скольких текстах из 140 встретилось данное слово); 3) употребительность. Причем коэффициент употребительности вычисляем по предложенной в работе М.В.Григорьевой (9) формуле (при равных подвыборках):

$$U_{jk} = \frac{n}{2} P (r-1),$$

где U_{jk} - индекс употребительности лексической единицы,
 n - количество подвыборок A_{τ} в составе A ,
 P - относительная частота,
 τ - показатель распространенности.

Наиболее употребительная лексика немецкого подъязыка экономики промышленности, отобранная нами согласно вышеуказанным принципам, составляет:

- а) частотный словарь немецких словоформ заданного подъязыка;
- б) частотный словарь немецких слов - понятий заданного подъязыка (на примере имен существительных);
- в) немецко-русский словарь наиболее употребительных имен существительных по данному подъязыку.

Предлагаемая методика отбора является лишь одной из попыток разрешить проблему отбора наиболее употребительной лексики подъязыка научно-технической специальности с помощью ЭВМ (на примере немецкого подъязыка ЭП).

Частотный словарь немецкого подъязыка экономики промышленности (ЧСПЭП) содержит 2400 лексем-существительных. В основном это терминология по экономике промышленности, а также термины родственных дисциплин, с которыми ЭП тесно связана (политэкономика, бухгалтер, финансы, статистика, экономика труда). В словарь включены 2400 имен существительных с частотой F , где $2 \leq F \leq 1147$, которые представляют интерес для анализа. Словарь пополнен рядом новых слов. Около 30% всех лексем, собранных в нем, отсутствуют в существующих отечественных немецко-русских словарях. В частности, в него вошли термины, связанные с научно-техническим прогрессом, с применением ЭВМ. Кроме того, переводы ряда терминов, зафиксированных ранее в нормативных немецко-русских словарях, а также в последнем немецко-русском экономическом словаре И.С.Елява и Л.Т.Багми, в данном словаре уточнены.

Слово-понятие	Род	F	n	U _{jk}	T
Betrieb	m	1147	137	77,9960	9
Erzeugnis	n	637	106	33,4425	9
Produktion	f	611	139	42,1590	9
Planung	f	531	108	28,4085	9
Entwicklung	f	508	116	29,2100	9
Aufgabe	f	426	140	29,6070	9
Arbeit	f	382	104	19,6730	9
Anwendung	f	255	83	10,4550	9
Industrie	f	248	74	9,0520	9
Grundlage	f	233	102	11,7665	9
Analyse	f	203	46	4,5675	9
Leitung	f	202	82	8,1820	9
Beispiel	f	193	92	8,7815	9
Technik	f	189	81	7,4400	9
Ergebnis	n	186	91	8,3700	9
Qualität	f	189	76	6,8625	9
Mittel	n	173	78	6,6605	9
Investition	f	168	46	3,7800	7
Prozent	n	165	45	3,6300	9
Kosten	pl	163	49	3,9120	9
Bedingung	f	156	89	6,8640	9
Prozess	m	154	109	8,3160	9
Erhöhung	f	152	81	6,0800	9
Methode	f	151	86	6,4175	9
Preis	m	151	66	4,9075	8
Ausnutzung	f	150	56	4,1250	9
Werkstätige	m	148	62	4,5140	8
Bedeutung	f	146	90	6,4970	9
Abteilung	f	145	27	1,7400	8
Durchführung	f	145	56	3,9875	9

ЛИТЕРАТУРА

1. Программы по немецкому языку для неязыковых специальностей высших учебных заведений. М., 1975, с.7.
2. Андреев Н.Д. Статистико-комбинаторные методы в прикладном языковедении. Л., 1967, с. 90-91.
3. Опельбаум Е.В. Научно-технические подъязкы и преподавание иностранных языков.-В сб.: Методика преподавания иностранных языков в вузе. М., 1976, вып. 107, с. 33.
4. Рейцак А.А. О принципах и методах изучения словарного состава. Тарту, 1966.
5. Мишев Р. Словари основной лексики. - В сб.: Методика преподавания иностранных языков за рубежом. М., 1967, с. 291.
6. Gougenheim G., Michea R., Rivonce R., Sauvageat A. Elaboration du francais élémentaire. Etude sur l'établissement d'un vocabulaire et d'un grammaire de base. P., 1956.
7. Шерва Л.В. Предисловие к кн.: Русско-французский словарь. 2-е изд. М., 1959, с. 5.
8. Андрющенко В.М. Частотные словари и их параметры. - ИЛШ, 1967, № 3, с. 34; Алексеев П.М. Статистическая лексикография. Л., 1975; Калинина Е.А. Изучение лексико-статистических закономерностей на основе вероятностной модели. - В сб.: Статистика текста. Л., 1968; Розанов Е.Д. Коэффициент стабильности и характеристика частотных списков. - В сб.: Вопросы лингвистики и автоматизации лингвистических работ. Тр.ЦНИИЯ. М., 1970, вып. 3.
9. Григорьева И.В. Количественный и качественный анализ немецкой лексики (подъязык технологии с языка). Канд.дис. Львов, 1974, с. 50-56.
10. Алексеев П.М. Частотные словари и приемы их составления. - В сб.: Статистика речи. Л., 1968, с. 61-63;
11. Засорина Л.Н. О статистике и автоматизации в лексикографии. - Филол.науки, 1963, № 4.
12. Морозенко В.В., Турук И.Ф. Лексический минимум по английскому языку для технических вузов. М., 1975, с. 3.
13. Зубов А.В., Лукьяненко К.Ф., Пиотровский Р.Г., Хотьшов Э.Н. Лексико-статистическое описание текста на электронно-вычислительных машинах. - В сб.: Статистика речи. Л., 1968, с.118-119.

14. Алексеев П.М., Лебедев В.М., Пустовойтова Л.А., Иблонская Н.И. Методика и техника составления двуязычных частотных словарей-минимумов. - В сб.: Инженерная лингвистика и оптимизация преподавания иностранных языков. М., 1976; Алексеев П.М. К методике составления переводных словарей-минимумов на основе одноязычных частотных словарей. - В сб.: Методы анализа текстов. Минск, 1975.
15. Лукьяненко К.Ф. Лексико-статистическое описание английского научно-технического текста с помощью электронно-вычислительной машины. Канд.дис. Минск, 1969; Статистика речи, с.108-119.
16. Вертель В.А., Вертель Е.В. Алгоритмы получения частотного словаря с учетом длины словоформ. - В сб.: Статистика текста. Минск, 1970, т. 2; Вертель В.А. Автоматический немецко-русский словарь для перевода научно-технических текстов (подязык экономической литературы, ГДР). - В сб.: Частные вопросы автоматического анализа текстов. Минск, 1972.
17. Алексеев П.М. Статистическая лексикология, с. 47; Андрюшенко В.М. Частотные словари и их параметры; Вишнякова Н.Г. Терминологическая лексика, методика и принципы ее сбора для обучения иностранному языку в технических вузах. Канд.дис. М., 1965, с. 66.
18. Динабург И.А. Принципы сбора терминологического минимума для неязыкового вуза. Канд.дис. Челябинск, 1962; Штеффальдт Э.А. Частотный словарь современного русского литературного языка. Таллин, 1963; Павин Л.М. Частотный англо-русский словарь общеэкономических терминов. Куйбышев, 1974, с. 11.
19. Каширина М.Б. Статистическое моделирование лексики научных текстов (подязык физики твердого тела). Канд.дис. Львов, 1973.
20. Итин Л.Г. Экономика социалистической промышленности. М., 1966; Справочник экономиста-плановика промышленного производства. Под ред. Н.И.Герзиско, А.И.Жданова. М., 1964; Омаров А.М. Экономика промышленного предприятия. М., 1974; Каменниер С.Б. и др. Организация и планирование промышленных предприятий. М., 1967.
21. Григорьева М.В. Количественный и качественный анализ немецкой лексики; Вишнякова Н.Г. Терминологическая лексика, методика и принципы ее сбора для обучения иностранному языку в технических вузах; Розанов Е.Д. Коэффициент стабильности и характеристика частотных списков.